JOURNAL OF MULTIVARIATE ANALYSIS 51, 178–200 (1994)

# New Perspectives on Linear Calibration

TATSUYA KUBOKAWA AND CHRISTIAN P. ROBERT

*University of Tokyo, Tokyo, Japan; and*
*URA CNRS 1378, Université de Rouen, Rouen, France*

In univariate calibration, two standard estimators are usually opposed: the classical estimator and the inverse regression estimator. Controversies have followed the use of both estimators and we consider them from a decision-theoretic perspective, establishing the inadmissibility of the classical estimator and the admissibility of the inverse regression estimator. The latter allowing for a Bayesian interpretation, we also develop a fully noninformative study of the calibration model and derive a reference prior which avoids the inconsistency drawbacks of the inverse regression estimator. © 1994 Academic Press, Inc.

## 1. INTRODUCTION

The calibration problem occurs in measurement settings where two measurement methods are available: one is extremely accurate but expensive (or time-consuming, destructive, etc.) while the other is less accurate but easier and faster. The functional relation between the two types of measurements is assessed through a *calibration* experiment, where the values of both measurements are known; this relation is then used in subsequent experiments to predict the value of the more precise measurement based on a sample of the more approximative measurement.

This setting is of major importance in physical and chemical measurements and we refer the reader to Rosenblatt and Spiegelman (1981) for a general discussion on the practical issues of calibration. See also Osborne (1991) who provides a detailed and recent survey of the statistical aspects of calibration estimation, including the different modelings and the competing estimation procedures involved.

In this paper, even though extensions to more complex settings are possible, we mainly consider the unidimensional linear calibration model, where the calibration experiment can be represented as

$$y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1, ..., n \tag{1.1}$$

Received April 6, 1993; revised December 1, 1993.

AMS 1990 subject classifications: 62N99, 62J99, 62F15, 62C10, 62C15.

Keywords and phrases: calibration problem, inverse regression, control problem, consistency, quadratic risk, admissibility, reference priors, nuisance parameters.



0047-259X/94 $6.00

and the prediction experiment as

$$y_{0j} = \alpha + \beta x_0 + \varepsilon_{0j}, \qquad j = 1, ..., k,$$

where the vectors $\beta$, $y_i$, and $y_{0j}$ belong to $\mathbb{R}^p$, the $x_i$'s are known values of the precise measurements and $x_0$ is the value to predict. We assume in addition that the error vectors $\varepsilon_i$ and $\varepsilon_{0j}$ are independent normal variables from $\mathcal{N}_p(0, \sigma^2 I_p)$ and that the common variance $\sigma^2$ is also unknown. Therefore, the aim of the experiment is to predict the quantity $x_0$ based upon $y_{01}, ..., y_{0k}$ given that $\alpha$ and $\beta$ have been estimated from (1.1).

This model is only a special case of the multivariate calibration problems of Brown (1982) but it can be easily extended to the case of a covariance matrix of the form $\sigma^2 V$ when $V$ is known, as in Lieftinck-Koeijers (1988). Oman (1991) considers a setup where the precise measurement is also random, which is called *comparative calibration* in the literature (see Osborne, 1991). The linear calibration model can be reduced without loss of generality to the estimation of a ratio of two normal means, as shown by Fieller (1954) who used this representation to derive confidence intervals (see also Hoadley, 1970). In fact, if $\hat{\beta}$ is the least square estimator of $\beta$, i.e.,

$$\hat{\beta} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})/c_x$$

for

$$\bar{x} = \sum_{i=1}^{n} x_i/n, \qquad \bar{y} = \sum_{i=1}^{n} y_i/n \qquad \text{and} \qquad c_x = \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

let $s_1$ and $s_2$ be the two sums of residual squared errors based on the calibration and the prediction experiments, respectively. Then, if $\bar{y}_0 = \sum_{j=1}^{k} y_{0j}/k$, the statistics $\hat{\beta}$, $\bar{y}$, $\bar{y}_0$ and $s = s_1 + s_2$ are minimal sufficient and mutually independent, distributed as

$$\hat{\beta} \sim \mathcal{N}_p\left(\beta, \frac{\sigma^2}{c_x} I_p\right),$$

$$\bar{y} \sim \mathcal{N}_p\left(\alpha + \beta\bar{x}, \frac{1}{n}\sigma^2 I_p\right),$$

$$\bar{y}_0 \sim \mathcal{N}_p\left(\alpha + \beta x_0, \frac{1}{k}\sigma^2 I_p\right),$$

$$s \sim \sigma^2 \chi^2_{(n-2)p+(k-1)p}.$$

Under the transformation group described by

$$\bar{y} \to \bar{y} + d, \qquad \bar{y}_0 \to \bar{y}_0 + d, \qquad \alpha \to \alpha + d,$$

for any real $d$, the estimation problem is invariant and we can reduce the model to

$$y \sim \mathcal{N}_p(\beta, \sigma^2 I_p), \qquad z \sim \mathcal{N}_p(x_0\beta, \sigma^2 I_p), \qquad s \sim \sigma^2\chi_q^2, \tag{1.2}$$

where $y = c_x^{1/2}\hat{\beta}$, $z = (\bar{y}_0 - \bar{y})(n^{-1} + k^{-1})^{-1/2}$ and $q = (n + k - 3)p$. The values $\beta$ and $x_0$ are reparameterized as $\beta$ and $x_0$ in (1.2) but actually correspond to $c_x^{1/2}\beta$ and $(x_0 - \bar{x})\, c_x^{-1/2}(n^{-1} + k^{-1})^{-1/2}$ in the original model. In this paper, we consider the reduced model (1.2) and estimate $x_0$ in (1.2) under the squared error loss

$$L(x_0, \delta) = (\delta - x_0)^2,$$

the corresponding quadratic risk being $R(\theta, \delta) = \mathbb{E}_\theta[(\delta(y, z, s) - x_0)^2]$ for $\theta = (\beta, x_0, \sigma^2)$.

Two estimators are usually proposed (and opposed) in this setup: the *classical estimator* of Eisenhart (1939) and the *inverse regression estimator* of Krutchkoff (1967), respectively

$$\delta_c(y, z, s) = \frac{y'z}{\|y\|^2} \qquad \text{and} \qquad \delta_I(y, z, s) = \frac{y'z}{\|y\|^2 + s}.$$

Note that $\delta_I(y, z, s)$ is an extension of Krutchkoff's (1967) estimator, but is different from what is usually called the *inverse regression* estimator, where $x_i$'s are regressed on $y_i$'s in a classical linear model like (1.1). (The usual inverse regression estimator in this setup is given in Appendix 1). See Osborne (1991) and Brown (1982) for a detailed account of the derivation of these estimators and of the controversies associated with them.

The classical estimator $\delta_c$ is the maximum likelihood estimator of $x_0$ for $p = 1$ and is consistent when $\beta \neq 0$. However, it behaves dramatically when $\beta$ gets close to 0 and has infinite quadratic risk when $p \leqslant 2$, although it is finite for $p \geqslant 3$ (see Lieftinck-Koeijers, 1988). On the contrary, the introduction of $s$ in the denominator of $\delta_I$ provides a stability $\delta_c$ is missing since the quadratic risk of $\delta_I$ is finite for $p + q > 2$ (i.e., $(n + k - 2)p > 2$ in the original parameters). As shown by Hoadley (1970), the inverse regression estimator can also be interpreted in a Bayesian way. However, despite Kruchkoff's (1967) first assertions, $\delta_I$ does not dominate $\delta_c$ uniformly for $p \geqslant 3$ under squared error loss. Moreover, $\delta_I$ suffers from a serious drawback, namely *inconsistency*, as pointed out by Berkson (1969). In fact, when $q$ goes to infinity, $s$ goes to infinity almost surely and $\delta_I$ goes to 0. The inverse regression estimator is also inconsistent in the original model, i.e., when $n$ goes to $+\infty$. Later studies (see Osborne, 1991) showed that $\delta_I$ and $\delta_c$ are not comparable and that the domination of $\delta_c$ by $\delta_I$ only occurs in a neighborhood of $x_0 = 0$.

In this paper, we first reconsider the comparison between the two usual estimators by establishing that $\delta_c$ is inadmissible (Section 2) while $\delta_I$ is admissible for the model (1.2) (Section 3). This property does not indicate a particular superiority of $\delta_I$, however, since its admissibility is a consequence of an "excessive" shrinkage towards 0 which prohibits domination when $x_0$ is close to 0. The second aim of the paper is thus to provide an alternative to both estimators, $\delta_c$ and $\delta_I$. To this effect, we analyze the model in Section 4 through a noninformative technique based on *reference priors*, introduced in Bernardo (1979), which differentiates between the parameter of interest, $x_0$, and nuisance parameters, $(\beta, \sigma^2)$. The resulting Bayes estimator is then consistent. Note that Liseo (1993) recently considered a similar approach to the unidimensional calibration model in term of confidence properties when $\sigma^2$ is known.

A side effect of the decision-theoretic developments of Section 2 is to point out a link between linear calibration problems and a *control problem*, which estimates a normal mean $\theta$ under the loss

$$L^c(\theta, \delta) = (\theta'\delta - 1)^2, \tag{1.3}$$

studied in Berger *et al.* (1982), Berliner (1983), and Srinivasan (1984). In fact, the domination of some estimators of $x_0$ under quadratic loss can be derived from the domination of corresponding estimators under the loss (1.3), even though there is no exact correspondence.

Some results in this paper can be extended to the case where $\varepsilon_i$, $\varepsilon_{0j}$ have a fully unknown covariance matrix $\Sigma$. The corresponding extension of the domination of the classical estimator is shortly treated in Appendix 1.

## 2. Inadmissibility of the Classical Estimator

We show in this section the inadmissibility of the classical estimator $\delta_c$ and propose some improved procedures for squared error loss, using inequalities similar to those of Kubokawa (1994).

When we consider estimators associated with nonnegative functions $g$, of the form

$$\delta_g(y, z, s) = g\left(\frac{\|y\|^2}{s}\right)\frac{y'z}{s},$$

including $\delta_c(g(t)=1/t)$ and $\delta_I(g(t)=1/(1+t))$, the risk of $\delta_g$ can be written as

$$
\begin{aligned}
R(\theta, \delta_g) &= \mathbb{E}_\theta\left[\left\{g\left(\frac{\|y\|^2}{s}\right)\frac{y'(z-x_0\beta)}{s}+g\left(\frac{\|y\|^2}{s}\right)\frac{y'\beta}{s}x_0-x_0\right\}^2\right]\bigg) \\
&= \mathbb{E}_\theta\left[g^2\left(\frac{\|y\|^2}{s}\right)\frac{\|y\|^2}{s^2}\sigma^2\right]+x_0^2\mathbb{E}_\theta\left[\left(g\left(\frac{\|y\|^2}{s}\right)\frac{y'\beta}{s}-1\right)^2\right]. \qquad (2.1)
\end{aligned}
$$

This implies that the domination of $\delta_g$ in the linear calibration problem is equivalent to finding a function $\tilde{g}$ which satisfies the two following conditons

$$
\mathbb{E}_\theta\left[\tilde{g}^2\left(\frac{\|y\|^2}{s}\right)\frac{\|y\|^2}{s^2}\right]\leqslant \mathbb{E}_\theta\left[g^2\left(\frac{\|y\|^2}{s}\right)\frac{\|y\|^2}{s^2}\right]; \qquad (\mathrm{P}_1)
$$

$$
\mathbb{E}_\theta\left[\left(\tilde{g}\left(\frac{\|y\|^2}{s}\right)\frac{y'\beta}{s}-1\right)^2\right]\leqslant \mathbb{E}_\theta\left[\left(g\left(\frac{\|y\|^2}{s}\right)\frac{y'\beta}{s}-1\right)^2\right]. \qquad (\mathrm{P}_2)
$$

uniformly in $\theta$, , since both expectations are independent of $x_0$. The problem $(\mathrm{P}_2)$ can be directly interpreted as a control problem with unknown variance, where $\beta$ is estimated under the loss (1.3). For an introduction to the control problem with known variance, see Berger *et al.* (1982) and the references therein.

In order to improve on $\delta_g$, we consider estimators of the form

$$
\delta_g^\varphi(y, z, s) = g\left(\frac{\|y\|^2}{s}\right)\left(1-\varphi\left(\frac{\|y\|^2}{s}\right)\right)\frac{y'z}{s} \qquad (2.2)
$$

with positive and absolutely continuous functions $g$ and $0<\varphi<1$. Note that condition $(\mathrm{P}_1)$ indeed implies that dominating estimators are necessarily *shrinkage estimators*. We assume without loss of generality that $R(\theta, \delta_g)<\infty$.

THEOREM 2.1. *Assume that*

(a) $\varphi(\omega)$ *is nonincreasing and* $\lim_{\omega\to\infty}\varphi(\omega)=0$,

(b) $\varphi(\omega)\leqslant\max\{0, \varphi_0(\omega)\}$, *where*

$$
\varphi_0(\omega)=\inf_{j\geqslant 0}\left\{1-\frac{2b-2+2j}{1+2j}\frac{\int_0^\omega g(z)\dfrac{z^{a+j}}{(1+z)^{b+j}}dz}{\int_0^\omega g^2(z)\dfrac{z^{a+j}}{(1+z)^{b+j-1}}dz}\right\} \qquad (2.3)
$$

*for* $a=p/2$ *and* $b=(p+q)/2$. *Then* $\delta_g^\varphi$ *given by* (2.2) *dominates* $\delta_g$.

The upper bound on $\varphi(\omega)$ given in condition (b) is rather intricated, since it involves an infimum on quantities depending on $g$. However, if $g(\omega)(1+\omega)$ is nonincreasing, using the inequality

$$\frac{\int_0^\omega g(z)\frac{z^{a+j}}{(1+z)^{b+j}}\,dz}{\int_0^\omega g^2(z)\frac{z^{a+j}}{(1+z)^{b+j-1}}\,dz} \leqslant \sup_{0<z<\omega}\left\{\frac{1}{g(z)(1+z)}\right\} = \frac{1}{g(\omega)(1+\omega)},$$

i.e.

$$\varphi_0(\omega) \geqslant 1 - \frac{2b-2}{g(\omega)(1+\omega)},$$

we can derive a simple solution satisfying the conditions (a) and (b), given by

$$\varphi^T(\omega) = \max\left(1 - \frac{q+p-2}{g(\omega)(1+\omega)}, 0\right).$$

Then we get an improved estimator of the form

$$\delta_g^{\varphi^T}(y, z, s) = \min\left(g\left(\frac{\|y\|^2}{s}\right), \frac{q+p-2}{1+\|y\|^2/s}\right)\frac{y'z}{s}. \tag{2.4}$$

Therefore,

COROLLARY 2.2. *Assume that* $p \geqslant 3$. *The classical estimator* $\delta_c$ *is inadmissible and dominated by*

$$\delta^T(y, z, s) = \min\left(\frac{1}{\|y\|^2}, \frac{q+p-2}{s+\|y\|^2}\right) y'z.$$

The above result directly extends to the cases $p = 1, 2$ since the risk of $\delta^T$ is then finite while $R(\omega, \delta_c) = +\infty$. Also, note that Theorem 2.1 does not apply for the estimator $\delta_I$ since condition (b) is then empty. In fact, the ratio of the two integrals in Theorem 2.1 is 1 when $g(t) = 1/(1+t)$ and $p+q-2+2j > 1+2j$ implies that the bound on $\varphi$ is negative. We show in Section 3 that $\delta_I$ is actually admissible. It is interesting to note that, for the control problem, the estimator $y/(\|y\|^2+s)$ is inadmissible for $p \geqslant 5$ (Zaman, 1981; Berliner, 1983). But the addition of condition (a) in Theorem 2.1 prohibits this result being used to prove domination.

*Proof.* By the argument following (2.1), we only need to consider problem ($P_2$). Making the transformations $x = y/\sigma$ and $v = s/\sigma^2$ and using the results of Berliner (1983, bottom of p. 817), we get

$$R^*(\beta, \sigma^2; \delta_g^{\varphi}) \equiv \mathbb{E}_{\beta, \sigma^2}\left[\left\{g\left(\frac{\|x\|^2}{v}\right)\left(1-\varphi\left(\frac{\|x\|^2}{v}\right)\right)\frac{x'}{v}\frac{\beta}{\sigma}-1\right\}^2\right]$$

$$= Z\left(\frac{\|\beta\|}{\sigma}\right)\int_0^{\|\beta\|/\sigma}\eta_1\left(\frac{\|\beta\|^2}{\sigma^2}-\eta_1^2\right)^{(p-3)/2}$$

$$\times\left[\int_0^{\infty}\int_0^{\infty}(-\mathscr{D}\psi(r))\, r^p(e^{r\eta_1}-e^{-r\eta_1})e^{-r^2/2}\,dr\, f_q(v)\,dv\right]d\eta_1,$$

where $f_q(v)$ denotes the density function of the $\chi_q^2$ distribution,

$$\psi(r) = g\left(\frac{r^2}{v}\right)\left(1-\varphi\left(\frac{r^2}{v}\right)\right)\frac{1}{v},$$

$$\mathscr{D}\psi(r) = \psi(r)\{2r\psi'(r) + (p+1-r^2)\,\psi(r) + 2\}$$

$$= g\left(\frac{r^2}{v}\right)\left(1-\varphi\left(\frac{r^2}{v}\right)\right)\frac{1}{v}$$

$$\times\left[\frac{4r^2}{v^2}\left\{g'\left(\frac{r^2}{v}\right)\left(1-\varphi\left(\frac{r^2}{v}\right)\right)-g\left(\frac{r^2}{v}\right)\varphi'\left(\frac{r^2}{v}\right)\right\}\right.$$

$$\left.+(p+1-r^2)\,g\left(\frac{r^2}{v}\right)\left(1-\varphi\left(\frac{r^2}{v}\right)\right)\frac{1}{v}+2\right]$$

and

$$Z(t) = 2\cdot(2)^{-p/2}(\pi)^{-1/2}\left[\Gamma\left(\frac{p-1}{2}\right)\right]^{-1} t^{2-p}\exp(-t^2/2).$$

Note that $\mathscr{D}\psi(r)$ is a function of $r^2$ and $v$ only, that is $\mathscr{D}\psi(r) = D(r^2, v)$; this implies

$$\int_0^{\infty}(-D(r^2, v))\, r^p[e^{r\eta_1}-e^{-r\eta_1}]\, e^{-r^2/2}dr$$

$$= \int_0^{\infty}(-D(x, v))\, f_{p+2}(x; \eta_1)\,dx,$$

where $f_{p+2}(x; \eta_1)$ is a mixture of central chi-squared densities,

$$f_{p+2}(x; \eta_1) = \sum_{j=0}^{\infty} p_j f_{p+2+2j}(x),$$

with

$$p_j = \frac{\eta_1^{2j+1}}{(2j+1)!} \Gamma\left(\frac{p+2+2j}{2}\right) 2^{(p+2+2j)/2}$$

and $f_{p+2+2j}(x)$ the density of $\chi^2_{p+2+2j}$. Hence,

$$\begin{aligned} & R^*(\beta, \sigma^2; \delta_g) - R^*(\beta, \sigma^2; \delta_g^\varphi) \\ & \quad = Z\left(\frac{\|\beta\|}{\sigma}\right) \int_0^{\|\beta\|/\sigma} \eta_1^2 \left(\frac{\|\beta\|^2}{\sigma^2} - \eta_1^2\right)^{(p-3)/2} \Delta(\eta_1; \delta_g, \delta_g^\varphi)\, d\eta_1, \end{aligned}$$

where

$$\begin{aligned} \Delta(\eta_1; \delta_g, \delta_g^\varphi) = \int_0^\infty \int_0^\infty & \left\{ g\left(\frac{x}{v}\right)\left[1 - \varphi\left(\frac{x}{v}\right)\right] \frac{1}{v} \left[\frac{4x}{v^2} \left\{ g'\left(\frac{x}{v}\right)\left(1 - \varphi\left(\frac{x}{v}\right)\right) \right.\right.\right. \\ & \left.\left. - g\left(\frac{x}{v}\right) \varphi'\left(\frac{x}{v}\right) \right\} + (p+1-x)\, g\left(\frac{x}{v}\right)\left(1 - \varphi\left(\frac{x}{v}\right)\right) \frac{1}{v} + 2 \right] \\ & \left. - g\left(\frac{x}{v}\right) \frac{1}{v} \left[\frac{4x}{v^2} g'\left(\frac{x}{v}\right) + (p+1-x)\, g\left(\frac{x}{v}\right) \frac{1}{v} + 2\right] \right\} \\ & \times f_{p+2}(x; \eta_1)\, f_q(v)\, dx\, dv. \end{aligned}$$

It is therefore sufficient to show that $\Delta(\eta_1; \delta_g, \delta_g^\varphi) \geqslant 0$ for all $\eta_1 \geqslant 0$.

For the sake of convenience, let us introduce

$$A(x, v) = g\left(\frac{x}{v}\right)\left\{\frac{4x}{v^3} g'\left(\frac{x}{v}\right) + \frac{p+1-x}{v^2} g\left(\frac{x}{v}\right)\right\},$$

$$B(x, v) = 2g^2\left(\frac{x}{v}\right) \frac{x}{v^3} \left(1 - \varphi\left(\frac{x}{v}\right)\right).$$

Since $\varphi(\infty) = 0$, using a definite integral argument as in Kubokawa (1994), we have

$$\begin{aligned} \Delta(\eta_1; \delta_g, \delta_g^\varphi) = \int_0^\infty \int_0^\infty & \left\{ \left(\varphi^2\left(\frac{x}{v}\right) - \varphi^2(\infty)\right) A(x, v) - 2\left(\varphi\left(\frac{x}{v}\right) - \varphi(\infty)\right) \right. \\ & \left. \times \left(A(x, v) + \frac{1}{v} g\left(\frac{x}{v}\right)\right) - 2\varphi'\left(\frac{x}{v}\right) B(x, v) \right\} \\ & \times f_{p+2}(x; \eta_1)\, f_q(v)\, dx\, dv \end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \int_0^\infty \left\{ -\int_1^\infty \frac{d}{dt}\left[ \varphi^2\left(t\frac{x}{v}\right) A(x,v) - 2\varphi\left(t\frac{x}{v}\right) \right.\right.\\
&\quad \left.\times \left( A(x,v) + \frac{1}{v} g\left(\frac{x}{v}\right)\right)\right] dt\\
&\quad \left. -2\varphi'\left(\frac{x}{v}\right) B(x,v) \right\} f_{p+2}(x;\eta_1)\, f_q(v)\, dx\, dv\\
&= 2\int_0^\infty \int_0^\infty \left\{ \int_1^\infty \varphi'\left(t\frac{x}{v}\right)\frac{x}{v}\left[ -\varphi\left(t\frac{x}{v}\right) A(x,v) \right.\right.\\
&\quad \left. + A(x,v) + \frac{1}{v} g\left(\frac{x}{v}\right)\right] dt\\
&\quad \left. -\varphi'\left(\frac{x}{v}\right) B(x,v) \right\} f_{p+2}(x;\eta_1)\, f_q(v)\, dx\, dv.
\end{aligned}
$$

Transforming $x$ to $y = tx$ and then transforming $t$ to $z = y/(vt)$ gives

$$
\begin{aligned}
&\int_0^\infty \int_1^\infty \varphi'\left(t\frac{x}{v}\right)\frac{x}{v}\left[ -\varphi\left(t\frac{x}{v}\right) A(x,v) \right.\\
&\qquad \left. + A(x,v) + \frac{1}{v} g\left(\frac{x}{v}\right)\right] f_{p+2}(x;\eta_1)\, dt\, dx\\
&\quad = \int_0^\infty \int_1^\infty \varphi'\left(\frac{y}{v}\right)\frac{y}{t^2 v}\left[ -\varphi\left(\frac{y}{v}\right) A\left(\frac{y}{t}, v\right) \right.\\
&\qquad \left. + A\left(\frac{y}{t}, v\right) + \frac{1}{v} g\left(\frac{y}{tv}\right)\right] f_{p+2}\left(\frac{y}{t};\eta_1\right) dt\, dy\\
&\quad = \int_0^\infty \int_0^{y/v} \varphi'\left(\frac{y}{v}\right)\left[ -\varphi\left(\frac{y}{v}\right) A(vz,v) \right.\\
&\qquad \left. + A(vz,v) + \frac{1}{v} g(z)\right] f_{p+2}(vz;\eta_1)\, dz\, dy.
\end{aligned}
$$

At last, transforming $y$ to $w = y/v$, we obtain

$$
\begin{aligned}
\Delta(\eta_1;\delta_g,\delta_g^\varphi) = 2\int_0^\infty \varphi'(w) \int_0^\infty v f_q(v)
&\times\left[ -\varphi(w)\int_0^w A(vz,v)\, f_{p+2}(vz;\eta_1)\, dz \right.\\
&+ \int_0^w \left( A(vz,v) + \frac{1}{v} g(z)\right) f_{p+2}(vz;\eta_1)\, dz\\
&\left. - B(vw,v)\, f_{p+2}(vw;\eta_1)\right] dv\, dw.
\end{aligned}
$$

Since $\varphi$ is nonincreasing, it is sufficient to show that

$$\int_0^\infty f_q(v)\left[(\varphi(w)-1)\left\{\int_0^w g(z)\left[\frac{4z}{v}g'(z)\right.\right.\right.$$
$$\left.+\frac{p+1-vz}{v}g(z)\right]f_{p+2}(vz;\eta_1)\,dz$$
$$\left.\left.-2g^2(w)\frac{w}{v}f_{p+2}(vw;\eta_1)\right\}-\int_0^w g(z)\,f_{p+2}(vz;\eta_1)\,dz\right]dv\geqslant 0$$

for all $\eta_1\geqslant 0$ and all $w\geqslant 0$ in the case of $\varphi'<0$. Hence, an integration by parts and the recurrence relation $xf_m(x)=(m-2)\,f_m(x)-f'(x)$ lead to

$$\begin{aligned}\int_0^w zg^2(z)\,f_{p+2+2j}(vz)\,dz &= -\frac{2w}{v}g^2(w)\,f_{p+2+2j}(vw)\\ &\quad+\frac{p+2+2j}{v}\int_0^w g^2(z)\,f_{p+2+2j}(vz)\,dz\\ &\quad+\frac{4}{v}\int_0^w zg(z)\,g'(z)\,f_{p+2+2j}(vz)\,dz,\end{aligned}$$

which is used to show that

$$\begin{aligned}&\int_0^w g(z)\left(\frac{4z}{v}g'(z)+\frac{p+1-vz}{v}g(z)\right)f_{p+2}(vz;\eta_1)\,dz\\ &\quad-2g^2(w)\frac{w}{v}f_{p+2}(vw;\eta_1)\\ &=\sum_{j=0}^\infty p_j\left\{\int_0^w g(z)\left(\frac{4z}{v}g'(z)+\frac{p+1-vz}{v}g(z)\right)f_{p+2+2j}(vz)\,dz\right.\\ &\quad\left.-2g^2(w)\frac{w}{v}f_{p+2+2j}(vw)\right\}\\ &=-\sum_{j=0}^\infty p_j\left\{\frac{1+2j}{v}\int_0^w g^2(z)\,f_{p+2+2j}(vz)\,dz\right\}.\end{aligned}$$

Hence $\Delta(\eta_1;\delta_g,\delta_g^\varphi)\geqslant 0$ for all $\eta_1>0$ if

$$\varphi(w)\leqslant 1-\frac{1}{1+2j}\frac{\int_0^\infty\int_0^w g(z)\,f_{p+2+2j}(vz)\,f_q(v)\,dz\,dv}{\int_0^\infty\int_0^w g^2(z)(1/v)\,f_{p+2+2j}(vz)\,f_q(v)\,dz\,dv}\tag{2.5}$$

for all $j \geqslant 0$ and all $w > 0$. Integrating out the numerator and the denominator in the r.h.s. of (2.5) with respect to $v$ yields condition (b) of Theorem 2.1, and the proof is thus complete. ∎

## 3. Admissibility of the Inverse Regression Estimator

As mentioned in the introduction, an alternative to the classical estimator $\delta_c$ is the *inverse regression estimator*

$$\delta_I(y, z, s) = \frac{y'z}{s + \|y\|^2}.$$

Although this estimator has been much criticized, in particular because of its inconsistency, the conditions of Theorem 2.1 do not apply to $\delta_I$, since the bound (b) is negative. This result is not surprising in the light of the following development, where we establish that $\delta_I$ is admissible for squared error loss, thus deriving an optimality property for the inverse regression estimator, when compared with the classical estimator. Hoadley (1970) (see also Dunsmore (1968), Halperin (1970) or Aitchinson and Dunsmore (1975)) pointed out the Bayesian nature of $\delta_I$.

We establish that $\delta_I$ is a proper Bayes estimator with finite Bayes risk for the reduced model (1.2), and therefore prove the admissibility of $\delta_I$ in (1.2) (see Berger, 1985). Introducing $\eta = 1/\sigma^2$, we consider the following class of proper prior distributions:

$$\beta \mid \eta \sim \mathcal{N}_p(0, 1/(t\eta) I_p) \quad (t > 0), \qquad \eta \mid x_0 \sim \mathcal{E}xp(\xi/2(1 + t + x_0^2)) \quad (\xi > 0),$$

$$x_0 \sim \mathcal{T}\left(p + q + 2, 0, \frac{1 + t}{p + q + 2}\right).$$

The joint density of $x_0$, $\eta$, $\beta$ is then

$$\pi(x_0, \eta, \beta) \propto \eta^{p/2}(1 + t + x_0^2)^{-(q+p)/2 - 1} e^{-\{t\|\beta\|^2 + \xi/(1 + t + x_0^2)\}\eta/2}. \quad (3.1)$$

Theorem 3.1. *The inverse regression estimator $\delta_I$ is proper Bayes for the prior distributions* (3.1) *and admissible under squared error loss for the model* (1.2) *if $p + q \geqslant 4$.*

*Proof.* The posterior distribution of $(x_0, \eta, \beta)$ satisfies

$$\pi(x_0, \eta, \beta \mid \mathcal{D}) \propto \eta^{(3p+q)/2}(1 + t + x_0^2)^{-(p+q)/2 - 1}$$
$$\times \exp\left\{-\left(t\|\beta\|^2 + \|y - \beta\|^2 + \|z - x_0\beta\|^2 + s + \frac{\xi}{1 + t + x_0^2}\right)\frac{\eta}{2}\right\},$$

where $\mathscr{D}$ denotes the data $(y, z, s)$. Therefore,

$$\pi(x_0, \eta \mid \mathscr{D}) \propto \eta^{q/2+p}(1+t+x_0^2)^{-p-q/2-1}$$
$$\times \exp\left\{-\left(\|y\|^2+\|z\|^2-\frac{\|y+x_0 z\|^2}{1+t+x_0^2}+\frac{\xi}{1+t+x_0^2}+s\right)\frac{\eta}{2}\right\},$$
$$\propto \eta^{q/2+p}(1+t+x_0^2)^{-p-q/2-1}$$
$$\times \exp\left\{-((\|y\|^2+s)(x_0-\delta_I)^2+D)\frac{\eta}{2(1+t+x_0^2)}\right\},$$

where

$$D = t\,\|y\|^2+(1+t)\,\|z\|^2+\xi+s(1+t)-\frac{(y'z)^2}{\|y\|^2+s}.$$

Integrating out the above term with respect to $\eta$ gives

$$\pi(x_0 \mid \mathscr{D}) \propto [(\|y\|^2+s)(x_0-\delta_I)^2+D]^{-p-q/2-1}.$$

This expression implies that the marginal posterior distribution of $x_0$ is a Student's $t$-distribution with mean $\delta_I(v, z, s)$, which establishes the fact that $\delta_I$ is a (proper) Bayes estimator.

To establish the admissibility of $\delta_I$, we show the finiteness of its Bayes risk. From (2.1), it follows that the risk function of $\delta_I$ is written by

$$R(\theta, \delta_I) = \mathbb{E}_\theta\left[\frac{\|y\|^2\sigma^2}{(s+\|y\|^2)^2}\right]+x_0^2\,\mathbb{E}_\theta\left[\frac{(y'(y-\beta)+s)^2}{(s+\|y\|^2)^2}\right]$$
$$= I_1 + x_0^2 I_2, \qquad \text{say.}$$

It is straightforward to check that, if $p+q \geqslant 3$, there exist constants $B_1$ and $B_2$ which are independent of the unknown parameter $\theta$ such that $I_1 \leqslant B_1$, $I_2 \leqslant B_2$. Since the prior distribution (3.1) is proper, the Bayes risk of $\delta_I$ is finite if

$$\int_{-\infty}^{\infty} \frac{x_0^2}{(1+t+x_0^2)^{(p+q)/2}}\, dx_0 < \infty,$$

which holds under the condition $p+q \geqslant 4$. The admisibility of $\delta_I$ is thus established. ∎

Hoadley (1970) considered instead the family of improper priors

$$\pi_H(x_0, \eta, \beta) \propto \eta^{p/2+1}(1+t+x_0^2)^{-(q+p)/2}\, e^{-t\,\|\beta\|^2\eta/2} \qquad (t>0) \quad (3.2)$$

and showed that $\delta_I$ is also generalized Bayes against these improper priors. Actually, $\delta_I$ is Bayes and generalized Bayes against a broad class of proper and improper prior distributions but the class (3.1) is more convenient to establish the admissibility of $\delta_I$.

However, while the above result definitely prevents the derivation of domination results similar to those of Section 2, it does not really justify the use of the inverse regression estimator. In particular, inconsistency still holds and is related to the prior (3.1) or (3.2). Indeed, one can see that the marginal prior on $x_0$,

$$(1 + t + x_0^2)^{-(p+q)/2}$$

gets more and more concentrated around 0 as $q$ goes to infinity. The limiting behavior of $\delta_I$ is then in accordance with this prior distribution. The next section is devoted to the derivation of an alternative Bayes estimator which escapes this undesirable behavior.

## 4. A Bayesian Alternative to the Inverse Regression Estimator

4.1. *Reference Priors.* Although $\delta_I$ is indeed a Bayes estimator for the family of priors given in (3.1) and (3.2), it can be argued that the Bayesian derivation of Hoadley (1970) was done a posteriori, namely after the ad hoc derivation of $\delta_I$ by *inverse regression arguments* in Krutchkoff (1967). The Bayesian justification of $\delta_I$ is therefore highly suspicious, especially when considering the results of Hunter and Lamboy (1981). In fact, they obtained quite different conclusions under a noninformative prior on $\beta$, $\eta = x_0\beta$ and $\sigma^2$, although this choice was itself much criticized by Hill (1981).

In this section, we propose to reconsider the Bayesian approach to the calibration problem by constructing a noninformative prior for this model based on the idea of reference prior. This approach was introduced in Bernardo (1979) and developed in Berger and Bernardo (1989, 1992a, 1992b). The reader is referred to these papers for additional references to a prior derivation technique which appears as an alternative to the usual Jeffreys method when nuisance parameters are present in the model. In addition, Ghosh and Mukerjee (1992) (see also Clark, 1989) have shown that the corresponding generalized priors have optimality properties in terms of asymptotic mean square error. Liseo (1993) recently proposed a reference prior in the inverse regression setup when $p = 1$ and $\sigma$ is known.

As for Jeffreys prior, the reference prior method stems from the Fisher information matrix. In this case, it is (see Appendix 1)

$$I(\beta, \sigma^2, x_0) = \begin{pmatrix} \frac{1+x_0^2}{\sigma^2} I_p & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & x_0\beta/\sigma^2 \\ 0 \cdots 0 & \frac{q+2p}{2\sigma^4} & \\ x_0\beta'/\sigma^2 & 0 & \|\beta\|^2/\sigma^2 \end{pmatrix}.$$

While the Jeffreys prior is given by $I(\beta, \sigma^2, x_0)^{1/2}$, i.e.,

$$\pi^J(\beta, x_0, \sigma^2) = (1+x_0^2)^{(p-1)/2} \|\beta\| (\sigma^2)^{-(p+3)/2},$$

(note the difference with (3.1) and (3.2), even when $t=0$), the algorithm proposed in Berger and Bernardo (1989) breaks the Fisher information $I(\beta, \sigma^2, x_0)$ in blocks corresponding to the nuisance parameters, $(\beta, \sigma^2)$, and the parameter of interest, $x_0$. It then follows from their algorithm (see Appendix 1) that

$$\pi_2(\beta, \sigma^2 | x_0) \propto |h_2|^{1/2}$$

where

$$h_2 = \begin{pmatrix} \frac{1+x_0^2}{\sigma^2} I_p & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0 & \sigma^{-4} \end{pmatrix}$$

and that $\pi_1(x_0)$ is proportional to $1/\sqrt{1+x_0^2}$, hence leading to the following result.

LEMMA 4.1. *The reference prior* $\pi^R(x_0, \beta, \sigma^2)$ *associated with the ordering* $(x_0, (\beta, \sigma^2))$ *is*

$$\pi^R(x_0, \beta, \sigma^2) = (\sigma^2)^{-(p+2)/2} \frac{1}{\sqrt{1+x_0^2}}.$$

Note that, similarly to the Jeffreys prior, this prior distribution does not involve $q$ contrary to the distributions (3.1) and (3.2). This is an advantage both from consistency (see below) and Bayesian points of view, since, as $q = (n+k-3)p$ in the original parametrization, this implies that the prior distribution does not depend on the sample size. We now consider the corresponding marginal posterior distribution of $x_0$, which can be derived explicitly. In fact,

$$\pi^R(\sigma^2, x_0, \beta \mid \mathscr{D}) \propto (\sigma^2)^{-(3p+q)/2-1} e^{-(s+\|y-\beta\|^2+\|z-x_0\beta\|^2)/2\sigma^2} \frac{1}{\sqrt{1+x_0^2}}. \tag{4.2}$$

Thus

$$\pi^R(\beta, x_0 \mid \mathscr{D}) \propto \left(s + \left\|\beta - \frac{y+x_0 z}{1+x_0^2}\right\|^2 \times (1+x_0^2) + \frac{\|x_0 y - z\|^2}{1+x_0^2}\right)^{-(3p+q)/2} \frac{1}{\sqrt{1+x_0^2}}$$

and

$$\pi^R(x_0 \mid \mathscr{D}) \propto (1+x_0^2)^{(p+q-1)/2} \left\{\left(x_0 - \frac{y'z}{s+\|y\|^2}\right)^2 + \frac{\|z\|^2+s}{\|y\|^2+s} - \frac{(y'z)^2}{(s+\|y\|^2)^2}\right\}^{-p-q/2}. \tag{4.3}$$

The posterior distribution of $x_0$ is therefore a modified version of the improper priors of Section 3 but the additional term $(1+x_0^2)^{(p+q-1)/2}$ moderates the shrinkage of $\delta_c$ towards $\delta_I$.

Since the (analytical) derivation of the Bayes estimator $\delta_R$ associated with $\pi^R$ is not possible, a theoretical comparison between $\delta_R$, $\delta_I$ and $\delta_c$ is not possible. We therefore conducted a Monte-Carlo simulation study to compare the risks of these estimators and of $\delta^T$ in the special case $\beta = (1, ..., 1)$, $\sigma = 1$ and $p = q = 5$. The computation of $\delta_R$ was done via Gibbs sampling according to an algorithm described in Appendix 2, since $\pi^R(x_0 \mid D)$ is only known up to a multiplicative factor. As shown by Fig. 4.1, the improvement brought by $\delta^T$ over $\delta_c$ is rather limited, while $\delta_I$ sees its risk rapidly increasing as $x_0$ drifts away from 0. That $\delta_R$ improves upon $\delta_I$ in the most of the parameter space is clearly exhibited by Fig. 4.1. The comparison with $\delta_c$ is more uncertain, although it seems there is domination of $\delta_c$ by $\delta_R$ for larger values of $x_0$.

4.2. *Consistency Issues.* The Bayes estimator associated with the posterior distribution (4.3) cannot be derived explicitly, except in special cases (hence the call to Markov Chain Monte-Carlo methods), and the study of its consistency is therefore more delicate. However, note that when

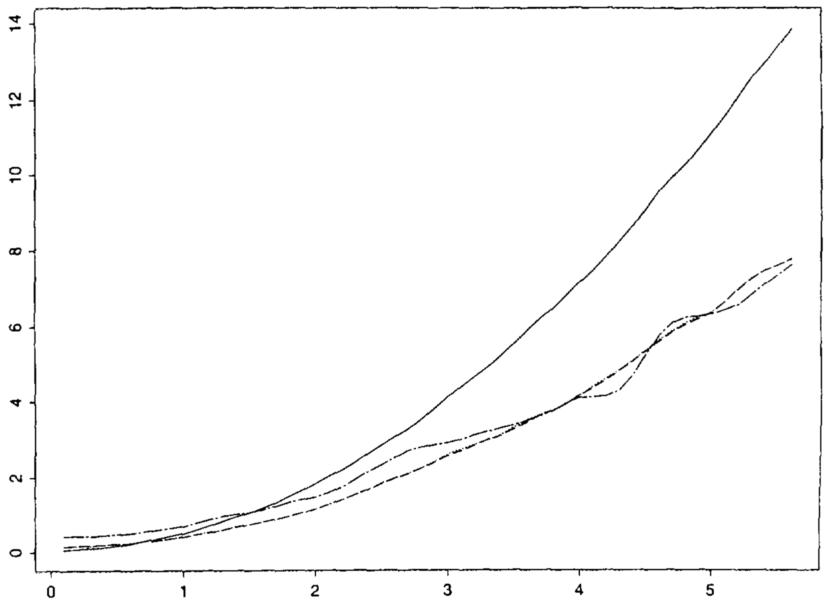


FIG. 4.1. Comparison of the risks of $\delta_I$ (plain), $\delta_c$ (dots), $\delta_T$ (dashes), and $\delta_R$ (dots and dashes) in the case $\beta = 1$, $\sigma = 1$, and $p = q = 5$.

$q$ goes to infinity, the distribution $\pi^R(x_0 \mid \mathscr{D})$ is increasingly concentrated around its mode. This posterior mode is solution of

$$(p+q-1)\frac{\delta}{1+\delta_R^2} - (2p+q)\frac{\delta_R-\delta_I}{\hat{\sigma}^2(2p+q+1)+(\delta_R-\delta_I)^2} = 0,$$

i.e.,

$$\delta_R^3(p+1) - \delta_R^2\,\delta_I(q-2) - \delta_R((\hat{\sigma}^2(2p+q+1)+\delta_I)(p+q-1) \\ -2p-q) - \delta_I(2p+q) = 0.$$

When $q$ goes to infinity, this equation gets transformed into

$$-\delta_R^3(p+1) + \delta_R^2\frac{y'z}{\sigma^2} + \delta_R\left(\frac{\|z\|^2-\|y\|^2}{\sigma^2} - p - 1\right) + \frac{y'z}{\sigma^2} = 0,$$

which can be approximated by

$$-\delta_R^3(p+1) + \delta_R^2\frac{x_0\,\|\beta\|^2}{\sigma^2} + \sigma_R\left[(x_0^2-1)\frac{\|\beta\|^2}{\sigma^2} - p - 1\right] + \frac{x_0\,\|\beta\|^2}{\sigma^2} = 0 \qquad (4.4)$$

when $\sigma^2$ is small. If, in addition, $x_0$ gets large, (4.4) can be replaced by

$$\delta_R^2 + x_0 \delta_R + 1 = 0,$$

with solution $(x_0 + \sqrt{x_0^2 - 4})/2$ which is equivalent to $x_0$ for larger values of $x_0$. These approximations show that the reference prior does not lead to the same undesirable behavior as the priors of (3.1) and (3.2), namely that the Bayes estimator associated with $\pi^R$ *is asymptotically* (*in* $x_0$) *consistent.*

## APPENDIX 1: An Extension to the Case of a Completely Unknown Covariance Matrix

We show here that the results of Section 2 can be extended to the case of a completely unknown covariance matrix $\Sigma$. When the error terms $\varepsilon_i$, $\varepsilon_{0j}$ in the original model have a $\mathcal{N}_p(0, \Sigma)$ distribution, the reduced model corresponding to (1.2) can be written as

$$y \sim \mathcal{N}_p(\beta, \Sigma), \qquad z \sim \mathcal{N}_p(x_0 \beta, \Sigma), \qquad S \sim \mathcal{W}_p(q, \Sigma).$$

The classical estimator of $x_0$ is then given by

$$\delta_c^* = \frac{y' S^{-1} z}{y' S^{-1} y}.$$

The usual inverse regression estimator arises from regressing the $x_i$ in (1.1) on the $y_i$. Brown (1982) showed that it has a form independent of the structure of the covariance matrix and is given by

$$\delta_I^* = \frac{y' S^{-1} z}{1 + y' S^{-1} y};$$

$\delta_I^*$ is thus different from $\delta_I$ given in Section 1 for $p \geqslant 2$. It may be noted that, for a completely unknown matrix $\Sigma$, $\delta_c^*$ and $\delta_I^*$ are related as $\delta_c$ and $\delta_I$ in the case $\Sigma = \sigma^2 I$.

We consider estimators of the general form $\delta_g^* = g(y' S^{-1} y)\, y' S^{-1} z$. The same argument as in (2.1) shows that

$$\begin{aligned} R(\theta, \delta_g^*) = {} & \mathbb{E}[g^2(y' S^{-1} y)\, y' S^{-1} \Sigma S^{-1} y] \\ & + x_0^2 \mathbb{E}[\{g(y' S^{-1} y)\, y' S^{-1} \beta - 1\}^2]. \end{aligned} \tag{A.1}$$

Let $u = \Sigma^{-1/2} y$ and $W = H \Sigma^{-1/2} S \Sigma^{-1/2} H'$ where $H$ is an orthogonal matrix such that $Hu = (\|u\|, 0, ..., 0)'$. Then $u$ and $W$ are independent,

$u \sim \mathcal{N}_p(\xi, I)$, $W \sim \mathcal{W}_p(q, I)$ for $\xi = \Sigma^{-1/2}\beta$. Let $W$ and $W^{-1}$ be partitioned as

$$W = \begin{pmatrix} w_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \qquad W^{-1} = \begin{pmatrix} w^{11} & W^{12} \\ W^{21} & W^{22} \end{pmatrix},$$

where $w_{11}$ and $w^{11}$ are scalar. By a usual matrix calculation, the second term in the r.h.s. of (A.1) can be represented as

$$x_0^2 \mathbb{E}[\{g(\|u\|^2 w^{11})(w^{11}, W^{12}) \|u\| H\xi - 1\}^2]. \tag{A.2}$$

Since $w^{11} = 1/w_{11.2}$, $W^{12} = -W_{12} W_{22}^{-1}/w_{11.2}$ for $w_{11.2} = w_{11} - W_{12} W_{22}^{-1} W_{21}$, and since, given $W_{22}$, $W_{12}$ has the conditional distribution $\mathcal{N}_{p-1}(0, W_{22})$, we get

$$\begin{aligned}
&\mathbb{E}\left[ g\left(\frac{\|u\|^2}{w_{11.2}}\right) \frac{\|u\|}{w_{11.2}} (1, -W_{12} W_{22}^{-1}) H\xi \right] \\
&= \mathbb{E}\left[ g\left(\frac{\|u\|}{w_{11.2}}\right) \frac{\|u\|}{w_{11.2}} (1, 0, ..., 0) H\xi \right] \\
&= \mathbb{E}\left[ g\left(\frac{\| \|^2}{w_{11.2}}\right) \frac{u'\xi}{w_{11.2}} \right]
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\mathbb{E}\left[ g^2\left(\frac{\|u\|^2}{w_{11.}}\right) \frac{\|u\|^2}{w_{11.2}^2} \xi' H' \begin{pmatrix} 1 & -W_{12} W_{22}^{-1} \\ -W_{22}^{-1} W'_{12} & W_{22}^{-1} W'_{12} W_{12} W_{22}^{-1} \end{pmatrix} H\xi \right] \\
&= \mathbb{E}\left[ g^2\left(\frac{\|u\|^2}{w_{11.2}}\right) \frac{\|u\|^2}{w_{11.2}^2} \xi' H' \begin{pmatrix} 1 & 0 \\ 0 & 1/(q-p) I_{p-1} \end{pmatrix} H\xi \right] \\
&= \mathbb{E}\left[ g^2\left(\frac{\|u\|^2}{w_{11.2}}\right) \frac{(u'\xi)^2}{w_{11.2}^2} \right] \\
&\quad + \frac{1}{q-p} \mathbb{E}\left[ g^2\left(\frac{\|u\|^2}{w_{11.2}}\right) \frac{\|u\|^2}{w_{11.2}^2} \xi' H' \begin{pmatrix} 0 & 0 \\ 0 & I_{p-1} \end{pmatrix} H\xi \right]
\end{aligned}$$

since $\mathbb{E}[W_{22}^{-1} W'_{12} W_{12} W_{22}^{-1}] = \mathbb{E}[W_{22}^{-1}] = 1/(q-p) I_{p-1}$. Hence (A.2) is expressed as

$$\begin{aligned}
&\frac{x_0^2}{q-p} \mathbb{E}\left[ g^2\left(\frac{\|u\|^2}{w_{11.2}}\right) \frac{\|u\|^2}{w_{11.2}^2} \xi' H' \begin{pmatrix} 0 & 0 \\ 0 & I_{p-1} \end{pmatrix} H\xi \right] \\
&+ x_0^2 \mathbb{E}\left[ \left\{ g\left(\frac{\|u\|^2}{w_{11.2}}\right) \frac{u'\xi}{w_{11.2}} - 1 \right\}^2 \right],
\end{aligned}$$

so that the domination of the estimator $g(y'S^{-1}y)\, y'S^{-1}z$ follows from Theorem 2.1. For instance, the classical estimator $\delta_c$ is improved upon by its truncated version

$$\delta^{*T} = \min\left\{\frac{1}{y'S^{-1}y}, \frac{q-1}{1+y'S^{-1}y}\right\} y'S^{-1}z.$$

Brown (1982) and Nishii and Krishnaiah (1988) treat a multivariate calibration model with a $p \times p$ coefficient matrix $\beta$ and $q$ explanatory variables $x_i$. An extension of Theorem 2.1 to this general model should be possible, but is not discussed here.

## APPENDIX 2: Fisher Information for the Calibration Model

For the model introduced in (1.2), the log-likelihood is given by

$$l(\beta, x_0, \sigma^2) = -(p+q/2)\log(\sigma^2) - \frac{s + \|y-\beta\|^2 + \|z - x_0\beta\|^2}{2\sigma^2},$$

up to an additive constant. Therefore,

$$\frac{\partial l}{\partial \beta}(\beta, x_0, \sigma^2) = -\frac{\beta - y + x_0(x_0\beta - z)}{\sigma^2}$$

$$\frac{\partial l}{\partial \sigma^2}(\beta, x_0, \sigma^2) = -\frac{2p+q}{2\sigma^2} + \frac{s + \|y-\beta\|^2 + \|z - x_0\beta\|^2}{2\sigma^4},$$

$$\frac{\partial l}{\partial x_0}(\beta, x_0, \sigma^2) = -\frac{\beta'(x_0\beta - z)}{\sigma^2}$$

and

$$\frac{\partial^2 l}{\partial \beta_i \partial \beta_j}(\beta, x_0, \sigma^2) = -\frac{1+x_0^2}{\sigma^2}\delta_{ij},$$

$$\frac{\partial^2 l}{\partial^2 \sigma^2}(\beta, x_0, \sigma^2) = \frac{2p+q}{2\sigma^4} - \frac{s + \|y-\beta\|^2 + \|z - x_0\beta\|^2}{\sigma^6},$$

$$\frac{\partial^2 l}{\partial^2 x_0}(\beta, x_0, \sigma^2) = -\|\beta\|^2/\sigma^2,$$

$$\frac{\partial^2 l}{\partial \beta_i \partial \sigma^2}(\beta, x_0, \sigma^2) = \frac{\beta_i - y_i + x_0(x_0\beta_i - z_i)}{\sigma^4},$$

$$\frac{\partial^2 l}{\partial \beta_i \partial x_0}(\beta, x_0, \sigma^2) = -\frac{2\beta_i x_0 - z_i}{\sigma^2},$$

$$\frac{\partial^2 l}{\partial x_0 \partial \sigma^2}(\beta, x_0, \sigma^2) = \frac{\beta'(x_0\beta - z)}{\sigma^4}.$$

This leads to the following information matrix for $(\beta, \sigma^2, x_0)$:

$$I(\beta, \sigma^2, x_0) = \begin{pmatrix} \frac{1+x_0^2}{\sigma^2} I_p & \begin{matrix} 0 \\ \cdots \\ 0 \end{matrix} & x_0\beta/\sigma^2 \\ 0 \cdots 0 & \frac{q+2p}{2\sigma^4} & 0 \\ x_0\beta'/\sigma^2 & 0 & \|\beta\|^2/\sigma^2 \end{pmatrix}.$$

The Jeffreys prior is then given by

$$\begin{aligned} \pi^J(\beta, \sigma^2, x_0) &= |I(\beta, \sigma^2, x_0)|^{1/2} \\ &= \left\{ \frac{2p+q}{2\sigma^4} \left( \frac{1+x_0^2}{\sigma^2} \right)^p \left| \frac{\|\beta\|^2}{\sigma^2} - \frac{x_0^2 \beta'\beta/\sigma^4}{(1+x_0^2)/\sigma^2} \right| \right\}^{1/2} \\ &= (1+x_0^2)^{(p-1)/2} \, \|\beta\| \, (\sigma^2)^{-(p-3)/2} \sqrt{\frac{2p+q}{2}}, \end{aligned}$$

according to the computation of the determinant of block matrices.

Similarly, the reference prior $\pi^R(\beta, \sigma^2, x_0)$ depends on $I(\beta, \sigma^2, x_0)$ in the following way (see Berger and Bernardo, 1989, for details): the conditional distribution $\pi^R(\beta, \sigma^2 | x_0)$ is proportional to $|h_2|^{1/2}$ where $h_2$ is the $(p+1, p+1)$ upper left corner of $I(\beta, \sigma^2, x_0)$. Therefore,

$$\begin{aligned} \pi^R(\beta, \sigma^2 | x_0) &\propto (1+x_0^2)^{p/2} (\sigma^2)^{-(p+2)/2} \\ &\propto (\sigma^2)^{-(p-2)/2}, \end{aligned}$$

which is independent of $x_0$. The "marginal distribution" $\pi^R(x_0)$ is then based on $|a_{11}|^{-1/2}$, if $a_{11}$ is the lower right $(1, 1)$ corner of $I^{-1}(\beta, \sigma^2, x_0)$, when eliminating the terms in $\beta$ and $\sigma^2$. In fact,

$$a_{11} = \frac{(1+x_0^2)^p (\sigma^2)^{-(p+2)}}{(1+x_0^2)^{p-1} \|\beta\|^2 (\sigma^2)^{-(p+3)}} \propto 1 + x_0^2$$

and

$$\pi^R(x_0) = \frac{1}{\sqrt{1+x_0^2}},$$

which actually leads to the reference prior proposed in Section 4.

## APPENDIX 3: GIBBS SAMPLING FOR THE REFERENCE PRIOR

A derivation of $\pi^R(x_0|\mathscr{D})$ given by (4.3) can be obtained by Monte-Carlo simulation methods like Gibbs sampling (see Gelfand and Smith (1990), Tierney (1991) or Casella and George (1992) for references). (Alternatives would be to use regular Monte-Carlo methods such as Importance Sampling or numerical integration.) Gibbs sampling can make use of the joint posterior distribution (4.2) and generate successively from the following three conditional distributions:

$$\sigma^2 \mid \beta, x_0 \sim (\sigma^2)^{-(3p+q)/2-1} e^{-(s+\|y-\beta\|^2+\|z-x_0\beta\|^2)/2\sigma^2},$$
$$\beta \mid x_0, \sigma^2 \sim e^{-\|(1+x_0^2)\beta-y-x_0z\|^2/2(1+x_0^2)\sigma^2},$$
$$x_0 \mid \beta, \sigma^2 \sim e^{-\|\beta x_0-z\|^2/2\sigma^2}/\sqrt{1+x_0^2}.$$

Moreover, since

$$\frac{1}{\sqrt{1+x_0^2}} \propto \int_0^{+\infty} e^{-(1+x_0^2)\tau}\tau^{-1/2}\, d\tau,$$

the above conditional distribution of $x_0$ can be decomposed into

$$x_0 \mid \beta, \sigma^2, \tau \sim e^{-\|\beta x_0-z\|^2/2\sigma^2} e^{-(1+x_0^2)\tau}$$
$$\tau \mid x_0 \sim e^{-(1+x_0^2)\tau}\tau^{-1/2},$$

this leading to a simple implementation of a Gibbs sampling algorithm for (4.3):

$$\sigma^2 \mid \beta, x_0, \tau \sim \mathscr{I}\mathscr{G}\left(\frac{3p+q}{2}, \frac{s+\|y-\beta\|^2+\|z-x_0\beta\|^2}{2}\right)$$
$$\beta \mid x_0, \sigma^2, \tau \sim \mathscr{N}_p\left(\frac{y+x_0z}{1+x_0^2}, \frac{\sigma^2}{1+x_0^2} I_p\right)$$
$$x_0 \mid \beta, \sigma^2, \tau \sim \mathscr{N}\left(\frac{\beta'z}{\|\beta\|^2+2\sigma^2\tau}, \frac{\sigma^2}{\|\beta\|^2+2\sigma^2\tau}\right)$$
$$\tau \mid x_0, \beta, \sigma^2 \sim \mathscr{G}a(1/2, 1+x_0^2).$$

## ACKNOWLEDGMENTS

This work was done while the first author was visiting LSTA, Université Pierre et Marie Curie, Paris. He gratefuly acknowledges a partial support from this department. Also the authors are grateful to Dr. S. Kuriki for his suggestions concerning relations between the linear calibration and the statistical control problems. Both referees went over the paper in great detail and contributed to a much improved revision.

## References

Aitchinson, J., and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge Univ. Press, London/New York.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

Berger, J. O., Berliner, L. M., and Zaman, A. (1982). General admissibility results for estimation in a control problem. *Ann. Statist.* **10** 838–856.

Berger, J. O., and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84** 200–207.

Berger, J. O., and Bernardo, J. M. (1992a). Ordered group reference priors with applications to the multinomial problem. *Biometrika* **79**, 25–37.

Berger, J. O., and Bernardo, J. M. (1992b). On the development of the reference prior method. In *Bayesian Statistics 4* (J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, Eds.). Oxford Univ. Press, London/New York.

Berkson, J. (1969). Estimation of a linear function for a calibration line: consideration of a recent proposal. *Technometrics* **11** 649–660.

Berliner, L. M. (1983). Improving on inadmissible estimators in control problems. *Ann. Statist.* **11** 814 826.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Royal Statist. Soc. Ser. B* **41** 113–147.

Brown, P. J. (1982). Multivariate calibration (with discussion). *J. Royal Statist. Soc. Ser. B* **49** 46–57.

Casella, G., and George, E. I. (1992). An introduction to Gibbs sampling. *Amer. Statist.* **46** 167–174.

Clark, B. (1989). Asymptotic cumulative risk and Bayes risk under entropy loss, with applications. Ph.D. dissertation, Univ. of Illinois.

Dunsmore, I. R. (1968). A Bayesian approach to calibration. *J. Royal Statist. Soc. Ser. B* **30** 396–405.

Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and statistical research. *Ann. Math. Statist.* **10** 162–184.

Fieller, E. C. (1954). Some problems in interval estimation. *J. Royal Statist. Soc. Ser. B* **16** 175–185.

Gelfand, A., and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

Ghosh, J. K., and Mukerjee, R. (1992). Non-informative prior. In *Bayesian Statistics 4* (J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, Eds.). Oxford Univ. Press, London/New York.

Halperin, M. (1970). On inverse estimation on linear regression. *Technometrics* **12** 727–736.

Hill, B. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics* **23** 335–338.

Hoadley, B. (1970). A Bayesian look at inverse linear regression. *J. Amer. Statist. Assoc.* **65** 356–369.

Hunter, W. G., and Lamboy, W. F. (1981). A Bayesian analysis of the linear calibration problem (with discussion). *Technometrics* **23** 323–328.

Kritchkoff, R. G. (1967). Classical and inverse regression methods of calibration. *Technometrics* **9** 429–439.

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Ann. Statist.*, to appear.

Lieftinck-Koiejers, C. A. J. (1988). Multivariate calibration: a generalization of the classical estimator. *J. Multivariate Anal.* **25** 31–44.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80** 295–304.

NISHII, R., AND KRISHNAIAH, P. R. (1988). On the moments of classical estimates of explanatory variables under a multivariate calibration model. *Sankhya Ser. A* **50** 137–148.

OMAN, S. D. (1991). Random calibration with many measurements: an application of Stein estimation. *Technometrics* **33** 187–195.

OSBORNE, C. (1991). Statistical calibration: a review. *Internat. Statist. Rev.* **59** 309–336.

ROSENBLATT, J. R., AND SPEIGELMAN, C. H. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics* **23** 329–333.

SRINIVASAN, C. (1984). A sharp necessary and sufficient condition for inadmissibility of estimators in a control problem. *Ann. Statist.* **12** 927–944.

TIERNEY, L. (1991). Exploring posterior distributions using Markov chains. In *Computer Science and Statistics: 23rd Symp. Interface*, pp. 563–570.

ZAMAN, A. (1981). A complete class theorem for the control problem, and further results on admissibility and inadmissibility. *Ann. Statist.* **9** 812–821.